SoPHIE: Social Robotic Platform for Human Interactive Experimentation

Target Application: Dialog Analysis by Bayesian Multimodal Observation on an Assistant Robot

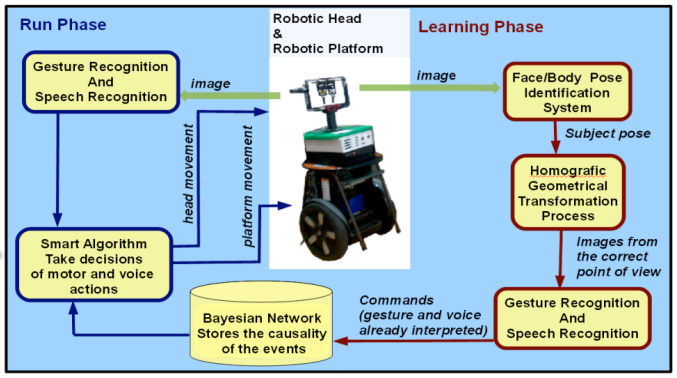
José Prado, Jorge Lobo and Jorge Dias

SoPHIE, the social robot:

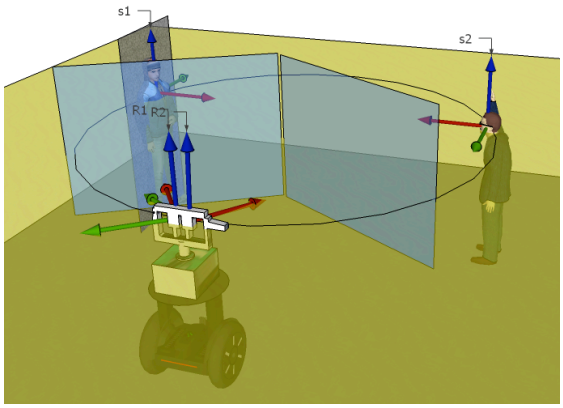
The aim of the *SoPHIE-project* is to develop a multipurpose platform to investigate social interaction between humans and robots. The objective of it's first version is to classify gestures and voice commands performed by two human actors performing a dialog (with a simplified language). SoPHIE vision uses a series of images taken by a stereo pair of cameras while it's audition is based on a pair of microphones.

In the figure aside a schema of the processes involved on the analysis algorithm. Each block then is explained briefly below.

Previous experimentation about Bayesian Learning were already done by the author at [3] , [4] , [5] and [6].



Homography Geometrical Transformation Process

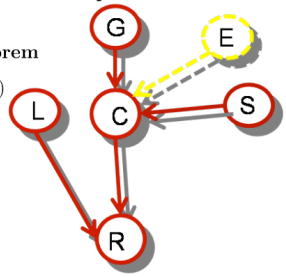


Our gesture recognition algorithm analysis the gestures based on the intrinsic parameters of the triangle that links head and hands [2]. So, what really matters is the proportional changes among the side's sizes and angles of it. Thus, we can say that the scale factor of the gesture triangle is irrelevant once guaranteed that the 3 corners (hands and face) are inside the camera field of view.
Considering this, and the figure aside, it is possible to remove the perspective projection. Then a virtual point of view, aligned to the other interlocutor is given to the robot.

Bayesian Network Stores the causality of the events

$P(C) = \text{Uniform}$
$P(R|C) = \text{acquired from learning by observation}$
$P(R) = \text{Uniform}$
$P(C|R) = \frac{P(C) \cdot P(R|C)}{P(R)}$

Bayesian Network



Based on Independence theorem

$P(L|C) = P(L) \& P(C|L) = P(C)$
Where:
G = gesture
S = sound (from human source)
E = emotion
L = laser, collision avoidance
C = Command
R = Response

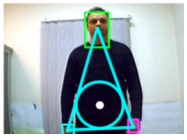
$$P(R|L,C)P(C|G,E,S)P(L)P(G)P(E)P(S)$$

Main Applications of Social Robots:

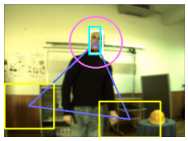
Security, Research, Health Care, Personal Assistant / Interface, toys, Business, Pet, Entertainment, Teacher, Transport, Companionship, Caregiver, Public Assistant

Face/Body Pose Identification System

Detecting face and hands



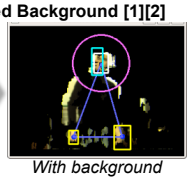
Without skin colour at the background



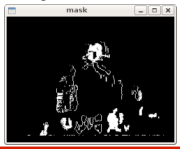
Problem occurs: With skin colour at the background

Solution – Horopter based Background [1][2]

Without background segmentation



With background segmentation

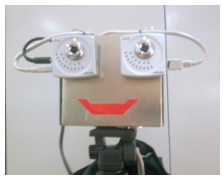


Gesture Recognition and Speech Recognition

In [5] we concluded that horopter is a valid approach for dynamic background segmentation, provided that it receives background with enough features, which usually happens. This segmentation enhances tracking results, both in speed and accuracy and should be further explored. Laban Movement Analysis is without a doubt a powerful movement descriptive tool, results show that it can, with some accuracy classify basic emotion primitives (contextualized within LMA), and the implementation of the remaining components is an ongoing work. To build an autonomous interactive multimodal social robot, we are here combining our current gesture classification algorithm [5] with the capability of auditory perception and dialog analysis.

Smart Algorithm take decisions of motor and voice actions

The main contributions of work [7] were the probabilistic description of the spoken dialogue process for simulation purposes. However, it was used a simple Bayesian Model that remains with static relationships between the variables.
Our proposal is to develop a Dynamic Bayesian Network (DBN). Description of work Initially it is necessary to correctly setup a tool for speech recognition. The selected tool is the Dragon Naturally Speaking software by Nuance. Our dialog analysis will be limited for one interlocutor speaking at a time, and also each interlocutor must be equipped with a microphone near to the mouth. We are also going to perform tests with Directional Microphones.



SoPHIE "downsized" version head

[1] [**Jose Prado**, Luis Santos and Jorge Dias] A Technique for Dynamic Background Segmentation using a Robotic Stereo Vision Head, in proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (**RO-MAN'09**) pag. 1035-1040, - Toyama, Japan, Sept. 27-Oct. 2, 2009

[2] [**Jose Prado**, Luis Santos and Jorge Dias] Horopter based Dynamic Background Segmentation applied to an Interactive Mobile Robot, 14th International Conference on Advanced Robotics (**ICAR'09**), Germany, from June 22th to 26th, 2009

[3] [**Jose Prado**, Jorge Lobo, Jorge Dias]Robotic Visual and Inertial Gaze Control using Human Learning, The 17-th International Conference on Computer Graphics, Visualization and Computer Vision, (**WSCG' 09**), Czech Republic, Fev. 2009

[4] [**Jose Prado** and Jorge Dias] Visuovestibular-Based Gaze Control Experimental Case, (**RECPAD' 08**) 14a Conferencia Portuguesa de Reconhecimento de Padroes, Coimbra, 2008

[5] [Luis Santos and **Jose Augusto** and Jorge Dias] Human Robot Interaction Studies on Laban Human Movement Analysis and Dynamic Background Segmentation, accepted on (**IROS' 09**), The 2009 IEEE/RSJ International Conference on Intelligent RObots and Systems, St Louis, USA

[6] [Jorge Lobo, Joao Filipe Ferreira, **Jose Prado**, Jorge Dias] Robotic Implementation of Biological Bayesian Models for Visuo-Inertial Image Stabilization and Gaze Control,(**IROS' 08**)

[7] Oliver Pietquin. A Framework for Unsupervised Learning of Dialogue Strategies.PhD thesis, Universitaires de Louvain, 2004

This work has been supported by EC-contract number P6-IST- 027140, Action line: Cognitive Systems.

BACS
Bayesian Approach to Cognitive Systems

Project: BACS - Bayesian Approach to Cognitive Systems http://www.bacs.ethz.ch/

Author e-mails: {jaugusto, jlobo, jorge}@isr.uc.pt

Mobile Robotics Laboratory
Institute of Systems and Robotics
ISR – Coimbra, Portugal